Exploring Logarithmic Models

Students investigate logarithmic relationships in demographic data about countries of the world, using an inquiry-based model, involving hypothesizing, experimental and computational modeling, and sense-making.

Lesson Goals	Students will be able to… • Model logarithmic relationships using functions
Student-facing Lesson Goals	• Let's use Pyret to model logarithmic relationships in data
Prerequisites	• Simple Data Types • Contracts: Making Tables and Displays • Exploring Linear Models • Row and Column Lookups • Exploring Quadratic Models • Exploring Exponential Models
Materials	• PDF of all Handouts and Page • Countries of the World Starter File • Fitting Wealth-v-Health (Desmos) • Exploring Logarithmic Functions (Desmos) • Fitting Wealth-v-Health, Part 2 (Desmos) • Lesson Slides • Printable Lesson Plan (a PDF of this web page)
Supplemental Materials	• Additional Printable Pages for Scaffolding and Practice

Key Points For The Facilitator	• This lesson assumes that students have been introduced to logarithms before, and are familiar with computation involving logarithmic functions.
	• The focus of this lesson is about *applications* of log functions, and builds an intuition for logarithmic computation by motivating substantial trial and error.

Glossary

asymptote :: A straight line to which a curve gets closer and closer - but never touches - as one of the variables approaches infinity.

form :: the shape of a relationship between two quantitative variables: whether the two variables together vary linearly or in some other way

logarithmic relationship :: A mathematical relation between two variables, x and y, in which a base must be raised to the power of y in order to equal the value of x. When graphed, a logarithmic relationship appears as a reflection of an exponential function across the diagonal line y = x.

Looking for Patterns   *30 minutes*

Overview

Students explore the [Countries of the World Starter File](#), and find a relationship between wealth and median life expectancy. They also learn how to build a new column for a table in Pyret.

Launch

- Do you think people living in wealthier countries generally live longer?
- Open the [Countries of the World Starter File](#)
- From the File menu, select "Save a Copy", and click "Run".
- Working in pairs or small groups, complete questions 1-6 on [Exploring the Countries Dataset](#).

Discuss students' findings. Pay special attention to the *form* that students use to describe the pattern in the scatter plot. Is it linear, or something else? Is it strong or weak? If there's some disagreement among the students, that's a good thing!

Be attentive to sense-making: If a wealthy country is suffering heavy casualties in a war, where would we expect to see that country's datapoint? Why? What other conditions would cause a point to shift up, down, left or right?

Investigate

Working in pairs or small groups, complete the rest of [Exploring the Countries Dataset](#) using [Fitting Wealth-v-Health (Desmos)](#).

Note: This activity involves lots of trial and error, as students try to come up with a good linear, quadratic, or exponential model. The goal is to discover that a new kind of model is necessary .

Synthesize

- What's the best model you came up with?
- What kinds of lines or curves did you draw?

Direct students to look at the scatter plot they created in Pyret of `median-lifespan` versus `pc-gdp` .

- How would you describe the shape of the scatter plot for `median-lifespan` versus `pc-gdp` ?
  - *It starts out rising fast on the left, before tapering off as* `pc-gdp` *increases to the right.*
- How well did our linear, quadratic, and exponential models fit this data?
  - *None of the kinds of functions we've seen so far exhibit the behavior we see here! In some ways, this behavior is the exact opposite of exponentials: instead of starting slow and taking off like a rocket, this curve explodes up the y-axis right away, before leveling off.*

Logarithmic Functions   *45 minutes*

Overview

Having identified that the wealth-v-lifespan scatter plot is neither linear, nor quadratic, nor exponential, students learn about characteristics of logarithmic functions in graphical, tabular, and function notation form.

Launch

We know that exponential functions grow by *equal factors over the same interval*, so instead of calculating "what is added to y", we calculate "what y is multiplied by" to produce the next value over a constant interval.

x_1	y_1
1	2
2	4
3	8
4	16
5	32

×2
×2
×2
×2

The behavior in this relationship is the *reverse* of exponential growth: instead of multiplying y by a constant factor to find the next value, we need to *multiply the size of our x-interval* by a constant factor just to get a constant growth in y! This relationship can be modeled using a *logarithmic function*.

x_1	y_1
2	1
4	2
8	3
16	4
32	5

×2
×2
×2
×2

- Complete <u>What Kind of Model? (Tables)</u>.
- What strategies did you use to decide if a table represented a quadratic, exponential, or logarithmic function?
- Was it especially difficult (or simple) to recognize any these function types? Why?

Have students share their answers. For logarithmic functions, it's especially important to have students talk about how much x needs to increase just to get a fixed increase in y. This foreshadows the idea of *base* for logarithmic relationships.

Logarithmic relationships - like the ones you identified on tables 1, 4, and 6 - grow or decay very quickly at first, but then level out as x increases. Our intervals need to grow longer and longer just to reach a small change in y. For example, it might take 10 steps to go from 1 to 2, but then a hundred more to get to 3 and a *thousand* more to get to 4.

Investigate

Logarithmic models have the form $f(x) = a \log_b(x) + c$

Review students answers, and then debrief via class discussion. Invite students to consider what new information they have gained by looking at graphical representations rather than tables.

The base b of the logarithm is similar to the base in exponential functions.

- 3^2 means "calculate 3 to the power of 2", which is 9.

- $\log_3(9)$ means "calculate what power 3 should be raised to in order to reach 9", which is 2.

Look back to your answers on [What Kind of Model? (Tables)](#). For tables that represent logarithmic functions, by what factor does x need to grow in order to get a constant increase in y?

logarithm	Translation	Solution
$\log_2(8)$	2 raised to what power is 8?	3
$\log_2(1)$	2 raised to what power is 1?	0
$\log_5(25)$	5 raised to what power is 25?	2
$\log_5(1)$	5 raised to what power is 1?	0
$\log_3(81)$	3 raised to what power is 81?	4
$\log_3(1)$	3 raised to what power is 1?	0

, (Don't forget — anything to the power of zero is always 1, so the log of 1 will always be zero - *for any base!*)

The term c is the vertical shift of the function, which moves the curve up or down.
(*We've seen vertical shifts in other kinds of functions given different names, like k for quadratics and b for linear function.*)

Because $\log_{anything}(1) = 0$

- the logarithmic term will always be zero at $x=1$

- y will always be the value of c at $x=1$

The term a is called the logarithmic coefficient, which - like b - determines how quickly the function grows. Extremely observant students may notice that there's a relationship between a and b, where the value of $2 \log_{10}(10) = \log_{10}(10^2)$!

Logarithmic models have a **vertical _asymptote_** where the function increases or decreases boundlessly. In this data exploration, the _asymptote_ will always be located on the y-axis ($x = 0$).

- We observed a logarithmic relationship between wealth and lifespan in our Countries of the World dataset.

- There are lots of relationships like this, where the function grows or decays very quickly at first, but then moves at a glacial pace the farther out it goes.

- Can you think of any other real-world relationships that logarithmic functions might do a good job modeling?

  - _Logarithms are primarily applied in science and technology. It may be a challenge for students to think of relevant examples. We've provided one for class discussion, below._

We measure sound intensity on a logarithmic scale, which proceeds in multiples of 10. The table to the right gives some intensity levels in _watts per square meter_ and in _decibels_. Our ears can hear incredibly quiet sounds (like a pin dropping), but also process incredibly loud sounds (like a fog horn). A fire alarm, for

Source	Intensity	Intensity level	× TOH
Threshold of hearing (TOH)	10^{-12}	**0 dB**	1
Whisper	10^{-10}	**20 dB**	10^2
Pianissimo	10^{-8}	**40 dB**	10^4
Normal conversation	10^{-6}	**60 dB**	10^6
Fortissimo	10^{-2}	**100 dB**	10^{10}
Threshold of pain	10	**130 dB**	10^{13}
Jet take-off	10^2	**140 dB**	10^{14}
Instant perforation of eardrum	10^4	**160 dB**	10^{16}

Table 1.1 from [Müller, FMP, Springer 2015]

example, is thousands of times louder than a dog barking, but it's difficult for our brains to process that much more "loudness". As a result, we also _perceive_ loudness on a logarithmic scale: for us to perceive a sound as being twice as loud as another, it actually has to be _a hundred times as loud_.

- Turn to [What Kind of Model? (Descriptions)](#) and practice identifying whether the scenarios are best modeled by linear quadratic, exponential, or logarithmic functions.

- What clues did you use to help you identify which relationships were which?

Have students share their answers. Be especially attentive to students who mis-label logarithmic relationships as "exponential" — the relationship between the two is extremely subtle!

Synthesize

- What *similarities* do you see between exponential and logarithmic functions?

- What *differences* do you see between exponential and logarithmic functions?

Changing the Scale   *30 minutes*

Overview

Students discover that when a logarithmic relationship is graphed on an *exponential* scale, the point cloud appears linear. When trying to use linear regression with those points, however, they are reminded that merely changing the scale of a graph does not actually change the data.

This section builds the foundation for *linearization*, transforming the points themselves, which students will do in the following section. (Note: this also opens the door for teaching inverse functions!)

Launch

Remember that logarithmic models have the form $f(x) = a \log_b(x) + c$

To fit our logarithmic model, we need to find a and c, such that the model fits the data as closely as possible.

- Complete the first section of Changing the Scale, using Fitting Wealth-v-Health, Part 2 (Desmos).
- What values did you come up with for a and c in your best-guess logarithmic model.
  - *Crowdsource and record students' responses on the board.*
- Were those values very similar or very different? What were the R^2 values?

Trial-and-error only gets us so far, and it's not clear that we would ever stumble upon the optimal model. **We need something like Pyret's `lr-plot` function, which uses computational methods to find the best possible model.** Unfortunately, `lr-plot` only finds linear models!

If only we could *transform* this data to make it appear linear. Then we could use `lr-plot` to fit the optimal model, and then reverse the transformation to get the optimal logarithmic model!

Investigate

Imagine that the scatter plot is printed on a sheet of rubber, and can be stretched or squashed in any way we want. Data Scientists often use transformations to stretch their data into shapes that are easier to use, and then reverse the transformation when they are done.

Complete the last section of Changing the Scale, using Fitting Wealth-v-Health, Part 2 (Desmos).

By *transforming the x-axis* to grow exponentially, we are "squashing" the coordinate plane so that each interval on the x-axis represents 10x the growth in `pc-gdp` as the one before it. This balances out the logarithmic growth in `median-lifespan`, makes the curved relationship appear linear, and warps our logarithmic model so it looks like a straight line-of-best-fit.

The following metaphor might help students make sense of this.

- A person running on a treadmill doesn't change location. Why not?
  - *Their forward movement is balanced by the backwards movement of the treadmill.*
- If they run faster and faster, what needs to happen to the treadmill to keep them in the same place?
  - *The treadmill needs to go faster as well. As long as the treadmill speed increases at the same rate as the runner, they will balance one another's growth.*
- How is the treadmill example comparable to what we've done with our x-axis transformation?

Going Deeper: Connecting to Inverse Functions

While this metaphor doesn't cover inverses in any real depth, the treadmill analogy opens the door to discussing how one kind of change can "cancel out" or "undo" another.

We are working to add additional material on inverse functions to our Algebra 2 materials, and hope to release them by Fall 2024!

Synthesize

- How does seeing the point cloud as linear help us think about logarithmic growth?
- Transforming the axes only makes things *look* linear - the actual points haven't changed at all, and we still can't use linear regression to find the best logarithmic model... Can you think of a way we could we transform the *data*, instead of the axes?

Transforming the Data   *45 minutes*

Overview

Having discovered that changing the scale of a graph does not allow them to use linear regression, students learn to transform the data into a linear shape, building a new column by applying a function to each row. This new data can be fit with a linear model. By then applying the *inverse* of this transformation to their computed linear model, they retrieve the logarithmic model.

Launch

We tried changing the scale on the x-axis from linear to exponential, which cancels out the logarithmic behavior by "shrinking the axis". Another strategy is to "shrink the data", by *transforming the x-coordinates themselves*. Instead of plotting `pc-gdp` on a **logarithmic** x-axis, we could plot `log(gdp)` on a **linear** x-axis.

> - Complete [Transforming the Data](#), using [Fitting Wealth-v-Health, Part 2 (Desmos)](#).
> - What values did you come up with for our coefficients m and b in your best-guess linear model.
>   - *Crowdsource and record students' responses on the board.*
> - Were those values very similar or very different?
> - How close were the values from our best-guess logarithmic model?
>   - *Students should discover that the values match the coefficients of the logarithmic model as well.*

Transforming the points instead of the axis has the same visual effect: the dots appear to fall in a straight line. But now we can plot them on a linear-scale axis, and use linear regression to find the best-possible model!

This transformation changed the *kind of growth* from logarithmic to linear: the term went from $a \log_{10}(x)$ to ax. Instead of increasing logarithmically by a, our new function increases *linearly* by a.

Investigate

We transformed the `pc-gdp` column in three steps:

1. We defined a transformation function, $g(x)$, which produces the log of whatever it's given (the `pc-gdp` column).

2. We defined a new column to use as our x-values, populating it by applying the transformation to each of our original x-values.

3. We displayed these transformed datapoints as a scatterplot.

It's easy to do the same thing in Pyret!

- Turn to [Logarithmic Models](#).
- Complete Part 1, then pause for class discussion.

Address any student questions about the Pyret function they've just discovered, `build-column.` Verify that students have recorded the slope and vertical shift for their regression line. Then, emphasize the key ideas below.

- At each point in our linear model, y is the *predicted median lifespan*, and x is the *log of per-capita gdp*.

- We want x to represent the **original, un-transformed** value, simply using *per-capita gdp* as-is...

- We've learned that:

  - the **slope** in the transformed, linear model is the same as the **log coefficient** in the un-transformed logarithmic model

  - the **vertical shift** in the transformed, linear model is the same as the **vertical shift** in the un-transformed logarithmic model

- Now complete Part 2 of [Logarithmic Models](#).

Just like in Desmos, transforming the `pc-gdp` column with a log function produces a scatter plot showing a linear pattern in the data! Pyret's `lr-plot` tool computes the best possible linear model for our transformed data, determining it to have a slope of `11.9011` and a y-intercept of `24.2636`. Our R^2 has jumped to 0.66311, showing a vastly better correlation than before.

y=11.9011x+24.2636; R: +0.81431; R-sqr: 0.66311

With the transformation applied, our linear model (in both function and Pyret notation) is:
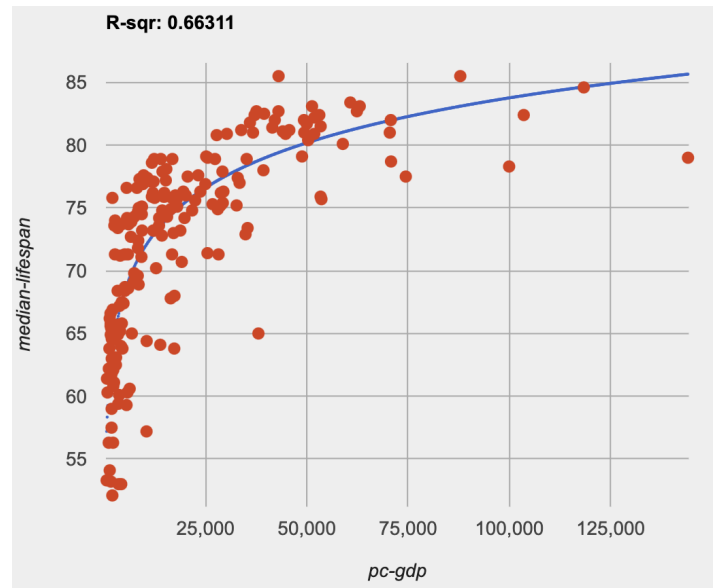
$$f(x) = 11.9011x + 24.2636$$
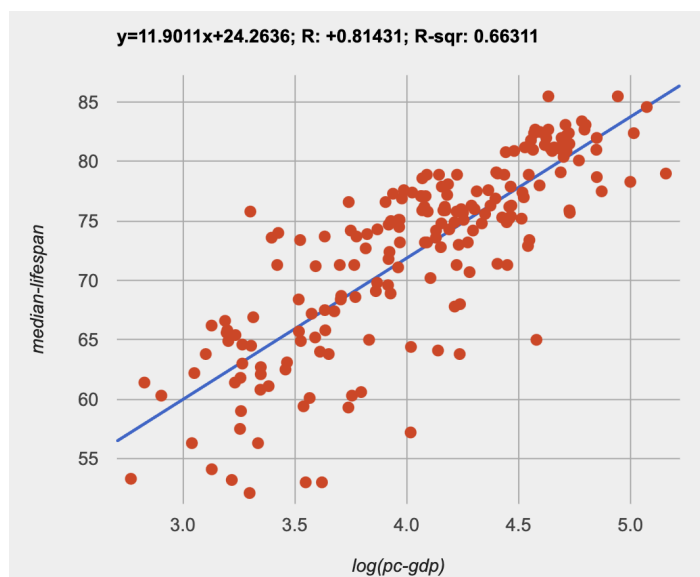
```
fun f(x): (11.9011 * x) + 24.2636 end
```

From [Transforming the Data](#), we know that the coefficients used in the transformed, linear model are the same ones used in the logarithmic, un-transformed model:

$$h(x) = 11.9011 \log_{10}(x) + 24.2636$$


```
fun h(x): (11.9011 * log(x)) + 24.2636 end
```

The resulting logarithmic model can be fit to our original scatter plot, showing a much better fit than our 2-point-derived estimates.

Synthesize

- Why is the R^2 value for our logarithmic model the same as the value for our linear model after transforming?
- Why were our coefficients for linear and logarithmic models the same, even though they were for different terms?
- How do you interpret this model?
- Why do you think the relationship between wealth and median lifespan is logarithmic?

Additional Exercises

For more practice transforming data and programming with filters:

Does Wealth impact lifespan equally if there's Universal Healthcare? is a guided activity that repeats the Data Science and Linearization techniques used here, but with the idea of exploring the relationship of universal healthcare with respect to wealth and median lifespan.

> ### Coming Soon!
>
> We are working on collecting more datasets that can be modeled with logarithmic functions so that we can offer students more practice with using linear regression to build logarithmic models.